DOCUMENT RESUME

ED 258 534 HE 018 492

AUTHOR Brum, Joseph; Ayers, Boyd
TITLE A Study to Determine the Effects of Two Different Methods of Evaluation on End of Course Grades in College Economics Classes. Theory and Practice of Adult Education Methodology.
PUB DATE 17 Aug 82
NOTE 23p.; Ed.D. Practicum, Nova University.
PUB TYPE Dissertations/Theses - Practicum Papers (043) -- Reports - Research/Technical (143)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS College Programs; *College Students; *Economics Education; *Educational Testing; *Evaluation Methods; *Grades (Scholastic); Higher Education; Low Achievement; *Student Evaluation; Technical Institutes; Two Year Colleges
IDENTIFIERS *Fayetteville Technical Institute NC

ABSTRACT

The effects of two methods of testing students on student grades were studied in 1982 with two economics classes that were taught by the same instructor. One class, which involved four 1-hour lectures weekly, received the normal four block-form examinations. The 34 students in this class were the control group. The 36 students in the other class received eight quizzes over smaller amounts of material. End of course grades, pre-test scores, and post-test results were compared at the end of the semester. The study was conducted because business and economics faculty were concerned about low student achievement in all economics classes at Fayetteville Technical Institute. Results showed that both evaluation methods were equally successful with respect to course grades. It is recommended that: course objectives be written; students' attitudes toward the evaluation methods used in economics classes be determined; and that the study be replicated. (Author/SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED258534

A STUDY TO DETERMINE THE EFFECTS OF TWO DIFFERENT METHODS OF EVALUATION ON END OF COURSE GRADES IN COLLEGE ECONOMICS CLASSES

THEORY AND PRACTICE OF ADULT EDUCATION METHODOLOGY

by

Joseph Brum, M. B. A.

Fayetteville Technical Institute

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Joseph Brum

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Dr. Boyd Ayers

Fort Bragg

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

A PRACTICUM PRESENTED TO NOVA UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF EDUCATION

NOVA UNIVERSITY

August 17, 1982

HE018 492

ABSTRACT

In the 1982 spring quarter, two economics classes were offered and taught by the same instructor. One class received the normal four block-form examinations, attending four one-hour lectures each week. The thirty-four (34) students in this class were designated the control group. The thirty-six (36) students in the other class, known as the experimental group, received eight (8) quizzes over smaller amounts of material. End of course grades, pre-test and post-test results were compared at the end of the semester using statistical techniques. The nature of the problem was that business and economics faculty members were concerned about low student achievement as measured by end of course grades in all economic classes at Fayetteville Technical Institute. The statement of the problem was as follows:

Do the students who receive eight (8) short quizzes over smaller amounts of material score significantly higher on end of course grades than do students who receive the normal four (4) comprehensive examinations?

Results of the study showed that both methods of evaluation were equally successful with respect to end of course grades. As a direct result of this study, the following recommendations were made:

(1) Replicate the study. This would provide an opportunity to further validate the findings.

(2) Write definite course objectives that clearly identify the material the instructor wants the student to learn.

(3) Determine students' attitudes toward evaluation methods used in economics classes.

Results clearly indicate that quizzes over smaller amounts of material are effective; however, future studies and experimentation in the area of test evaluation in the area of test evaluation will be required.

TABLE OF CONTENTS

	Page
INTRODUCTION	1
STATEMENT OF THE PROBLEM	3
HYPOTHESIS	4
BACKGROUND AND SIGNIFICANCE OF THE STUDY	4
DEFINITION OF TERMS	6
LIMITATIONS OF THE STUDY	7
BASIC ASSUMPTIONS	7
PROCEDURES FOR COLLECTING DATA	8
PROCEDURES FOR TREATING DATA	11
RESULTS AND FINDINGS	12
DISCUSSION, IMPLICATIONS, AND RECOMMENDATIONS	14
REFERENCES	17

INTRODUCTION

The problem of this study was to investigate and compare the effects of two methods of evaluation used in economics classes at Fayetteville Technical Institute, and to determine what effect these methods had on end of course grades, if any. Low student achievement as measured by end of course grades has been an area of major concern for the faculty for the past several years. This study attempted to show what relationships exist between different evaluation procedures and this problem.

Cohen (1969) suggests that student failure is really school and teacher failure in that failure is the result of vagueness in instructional goals rather than weakness in student ability. Hutchins (1968) speaks of the same problem as "educational shock." Perhaps the economics staff needs to reassess what that first major examination over three or four weeks of material does to students. It may be that the "educational shock" is too great for many and ultimately leads to a withdrawal or a poor grade in the course.

The "quiz system" in economics, as it is referred to, is really a trade-off on part of the "Mink Model" for learning (Mink, 1970). The "Mink Model" is based on the principle of reinforcement learning. It includes four basic components of learning theory:

(1) Predisposition for learring.

(2) Structure.

(3) Sequence.

(4) Reward.

Cross (1971) states that fear of failure will be the major impediment for the "new" students of the near future. The "new" students will be those who rank in the lowest third on tests of academic aptitude, and it is the group which constitutes the greatest reservoir of attenders.

The "quiz system" system seems to be supported by what B. F. Skinner says, "Man is a biological organism reacting to stimuli in his environment. He is a product of his experiences, his past learning or conditioning, and of potential reconditioning." Skinner sees man's behavior as predominantly influenced by feelings, frustrations or satisfaction (Wilson, 1969).

Roueche and Pitman (1972) make the following observation:

> Learning in its natural state proceeds from a response. A small child hears sound, makes a noise (response), and is reinforced. If he is punished, he is less likely to make further sounds. If he is ignored, he also begins to become more quiet. In short, a person does something and is then directed by the feedback.

The change in the evaluation system means ultimately that the student gets more feedback and, most likely, positive feedback. If this occurs, he goes on and finally succeeds in finishing the course successfully.

Finally, the tests used in economics since the course became a large team-taught course in many instances seem to have one primary purpose: To sort students into letter grade

categories. Kryspin and Feldhusen (1974) suggest that tests have a multiple purpose:

(1) Diagnostic -- highlighting student strengths and weaknesses.

(2) Motivational -- students are prompted to succeed on tests.

(3) Self-Evaluation -- students learn to develop their own self-evaluation skills.

(4) Instructional -- provides for a review of course material.

(5) Improving Teacher Effectiveness -- indicates to the teacher how well material was covered.

Testing can have an educational value in itself, how it can reveal learning difficulties, and -- particularly important -- how it can help improve teaching. Achievement tests, for example, are designed to aid in academic guidance, curriculum planning, and decisions on admissions, as well as to assess the individual's progress (Chauncey, 1963).

STATEMENT OF THE PROBLEM

For the purpose of this practicum, the research question that was examined was as follows: Of all students enrolled in basic economics classes, do the students who receive eight (8) short quizzes score significantly higher on end of course grades than do students who receive four (4) comprehensive examinations?

HYPOTHESIS

In order to properly examine the research problem, a hypothesis was formulated. The hypothesis for this problem was: Students who received eight (8) short quizzes score significantly higher on end of course grades than do students who receive four (4) comprehensive examinations.

BACKGROUND AND SIGNIFICANCE OF THE STUDY

Practicing educators who are concerned about the academic well being of their students are well aware that one of the major problems encountered by a sizeable proportion of them, both successful and unsuccessful, is that of test anxiety -- an apprehension about the experience of being assessed. Indeed, few people would deny a degree of apprehension about the testing situations encountered in everyday life -- school examinations, driving license tests, et cetera. Sufficient information is known about test anxiety and its relationship to human behavior to enable us to recognize that the parameters of the issue are numerous, and that the relationships amongst these parameters are complex. Indeed, the phenomenon of test anxiety has well been described by some researchers as a veritable "can of worms" (Fearn-Wannan, 1980).

In the present day, particular interest is centered upon the emotional state of anxiety in test situations, both as a consequence of the student's interactions with the educational environment, and as itself an influence upon

academic performance. The students' perceptions of their educational environment and their personality characteristics are linked with test anxiety and academic performance in a casual model, the parameters of which have a sound empirical basis (Fearn-Wannan, 1977; 1979).

There is little research-based information about current testing practice. Approximately twelve years ago, Kirkland (1970) reviewed the literature on test impact on students and schools and found that while much had been written about tests, few empirical studies were evident. In her review, Kirkland excluded the issues of test validity and misuse. She states that "Many of the issues and related questions are unanswerable by empirical means; for other issues, there is only indirect evidence." What is significant about her exclusions is the correct observation that these issues are "implications", often not found in empirical research. Today, there still remains a plethora of publications on these very issues and a dearth of empirical support on actual test practices.

Kirkland's review of the literature is concentrated mainly upon the social and psychological issues in testing (e.g., self-concept, anxiety, motivation, level of aspiration), more than upon instructional issues (e.g., curriculum, evaluation and pedagogy). Also, then as now, little empirical research had accumulated on the latter.

There is some evidence that curriculum-embedded tests and teacher-made tests matter in the course of instructional

decision making. In a study of California schools, Yeh (1978) reports that fifty-five percent (55%) of the teachers surveyed indicated that they regularly construct their own tests, which is an indication that some tests, if not published ones, are viewed by teachers as useful for some purpose. She also found that for assessing student progress, teachers tend to rely on more informal mechanisms such as observation and interactions with students than on the results of any tests. Others have only suggested teacher-made assessments as alternatives to standardized testing and offer techniques for teachers to help them make and use tests appropriately (Baron, 1958; Bauerfeind, 1963; Gorrow, 1966; Traxler, 1963; Quinto, 1977).

DEFINITION OF TERMS

To facilitate reading the remainder of this practicum, terms are defined below that are either unique, or are operationalized for this study.

Economics - A science designed to explain the nature of the real world by keeping the integration of theory and application always before the student.

Non-Major - Students with academic majors outside the area of economics.

Traditional Economics Course - A teacher-oriented course that consists of four one-hour lectures each week. Traditional forms of instruction are defined as instruction

characterized by the use of lectures with the opportunity for students to ask questions.

Withdrawal (W - grade) - Grade issued when a student drops a course if the student is passing at the time of the drop.

LIMITATIONS OF THE STUDY

(1) This study is limited to a comparison of only two classes: one economics class received four (4) comprehensive examinations versus the second class, which received eight (8) short quizzes over smaller amounts of material.

(2) Only end of course grades were compared in the two classes.

(3) No attempt was made to determine what the reasons were for withdrawal.

(4) No analysis was made of the differences in student environmental background.

(5) Implications gained through information attained from this study may only have value to the Business Administration/Economics staff at Fayetteville Technical Institute.

BASIC ASSUMPTIONS

(1) The two classes, one economics class received four (4) comprehensive examinations and the second class, which received eight (8) short quizzes over smaller amounts of material, studied are reasonably homogeneous.

(2) The same teacher for both classes provided equal learning opportunities for all students.

(3) The two classes involved in this study each had a normal distribution of intelligence.

(4) The students in both classes had equal opportunity to achieve certain grades.

(5) The pre-test did not sensitize the students to the post-test.

PROCEDURES FOR COLLECTING DATA

For this study, a classroom experiment strategy was used to facilitate the evaluation of the effectiveness of short quiz testing as it would occur in an actual classroom setting. The experiment was conducted in the Department of Business and Economics at Fayetteville Technical Institute, a two-year community college located in Fayetteville, North Carolina.

In the 1982 spring quarter, two economics classes were offered and taught by the same instructor. One class received the normal four (4) comprehensive examinations after attending four (4) one-hour classes each week. The thirty-four (34) students in this class were known as the control group. Thirty-six (36) students in the other group were known as the experimental group. This group was administered eight (8) short quizzes, over smaller amounts of material, after attending four (4) one-hour classes each week. The experimental and control groups received

a pre-test at the beginning of the experiment and a post-test at the conclusion. The mean scores on the pre-test and post-test results for matching students were compared. At the conclusion of the experiment, a comparison was made of the end of course grades to determine if there was any significant difference between the groups.

Subjects for the study were not randomly selected on an individual basis, but were selected as a group. An assumption was made that students in both groups were characteristic of a representative sampling of community college students.

The experimental group consisted initially of thirty-six (36) students (seventeen (17) females and nineteen (19) males), while the control group consisted of thirty-four (34) students (thirteen (13) females and twenty-one (21) males). The mean age of the experimental group was 23.6 years, while it was 26.5 years for the control group. However, at the conclusion of the experiment, useable data was obtained from only thirty (30) students in each of the experimental and control groups. Since matched test scores were necessary to conduct a student t-test, not all participants received both the pre-test and post-test, resulting in the aforementioned figures.

Material used in the lectures for both groups was taken primarily from the course textbook, Basic Economics, by Edwin R. Dolan; 1980; The Dryden Press. Although some

supplemental references were utilized, the majority of the material was taken from the textbook.

During the first instructional period, students in both the experimental and control groups were administered a pre-test. The pre-test consisted of fifty (50) items, to include twenty-five (25) multiple choice and twenty-five (25) true and false statements. The same statements were rearranged and administered as a post-test at the conclusion of the experiment. In preparing the test that had been utilized as pre-test and post-test, questions were selected from the instructor's manual, which accompanied the course textbook. Questions were selected which were representative of the material covered throughout the experiment. An assumption was made that the measurement instrument (pre-test and post-test), which would be used to test the students, was based on an interval means of measurement. An interval means of measurement was essential since this study was intended to analyze the eventual end of course grades utilizing the student t-test.

After the pre-test was administered, the student answer sheets were graded and the results retained for later analysis. It should be noted that students used IBM mechanically graded answer sheets during the pre-test and post-test. Since the grading of these tests was mechanical, student answer sheets were spot checked to insure correct grading by the machine.

The pre-test results were not shown to the students until after the completion of the experiment. These procedures were adopted for several reasons. First, in adherence with principle of "ceteris paribus", it was felt that to inform the students of their pre-test results may create additional factors which could influence the final results. For example, a student who was informed that he received a very low pre-test score may be motivated to a high degree to improve his performance. If he were to score high on the post-test, his performance would be attributed to his increased motivation and not as a result of receiving short, periodical quizzes. Undoubtedly, such a situation would create a dilemna when interpreting the results of the study.

During the last instructional period, both groups were administered a post-test. As noted earlier, the post-test was nothing more than a rearrangement of the questions utilized in the pre-test. Also, it was assumed that maturation process would not affect the post-test scores. Since the students again utilized a machine-graded IBM answer sheet, the completed answer sheets were spot checked to insure correct grading by the machine.

PROCEDURES FOR TREATING DATA

Once the data had been assembled and differences in end of course grades noted, the grades were submitted to a t-test to determine the significance between the means of

the samples. The null hypothesis tested was: There is no significant difference between the end of course grades of students receiving eight (8) short quizzes and those students who receive four (4) comprehensive examinations. The hypothesis was tested at the .05 level of significance, utilizing the degrees of freedom, which were 58, H_0: E = C. The alternative hypothesis is: H_a: E ≠ C.

RESULTS AND FINDINGS

The random sampling technique that was used in the investigation included a total of sixty (60) students that were studied. The experimental group consisted of twenty-three (23) freshmen and thirteen (13) sophomores, of which nineteen (19) students were males and the remaining seventeen (17) students were females. The control group consisted of nineteen (19) freshmen and fifteen (15) sophomores with twenty-one (21) of the students being males and the remaining thirteen (13) students females.

The group that received short quizzes was found to have a mean score of 2.53 and a standard deviation of 1.26. The group that received comprehensive examinations was found to have a mean score of 2.43 and a standard deviation of 1.02. This showed a mean score difference of .10.

An analysis of the data was then performed to determine if this difference was significant. An independent t-test for testing the significance of the difference between two independent sample means was used for the analysis.

The desired level of significance was established at .05. The degrees of freedom were calculated. Since the experimental group contained thirty (30) scores, and the control group also contained thirty (30) scores, the degrees of freedom were calculated to be 58. Next, the critical t value was established. By referring to a table of critical t values, it was found that for 58 degrees of freedom, the critical t value at the .05 level of significance was 2.00 for a two-tailed test. The standard procedure for t-testing was then employed. When the calculations were completed, the calculated t value was found to be .3536. Since the calculated value of t did not exceed the critical value of t, the null hypothesis that there is no significant difference between the end of course grades of students receiving short quizzes and those students who receive comprehensive examinations failed to be rejected. Final end of course grades were as follows:

CLASS	NUMBER	4	3	2	1	0	W
Experimental	36	7	15	7	1	0	6
Percent		19%	42%	19%	3%	0%	16%
Control	34	6	12	10	2	0	4
Percent		18%	35%	29%	6%	0%	12%

This study revealed that experiences in mean scores on the pre-tests and post-tests between the experimental and control groups in the 1982 spring quarter were not significant

at the .05 level of significance. In the pre-test exercise, the experimental group received a mean score of 47.66, and a standard deviation of 8.70. This was compared to the control group's pre-test results of 45.20 (mean score) and a standard deviation of 7.70. The t-test difference value was 1.29. In the post-test results, the experimental group received a mean score of 80.19 and a standard deviation of 10.57, as compared to a mean score of 82.25 and a standard deviation of 9.19 for the control group. This resulted in a t-test difference value of 1.15.

DISCUSSION, IMPLICATIONS, AND RECOMMENDATIONS

An analysis of the data obtained from this investigation failed to reject the null hypothesis. This indicates that students who received eight (8) short quizzes will not necessarily receive a higher end of course grade than a student who has received four (4) comprehensive examinations.

As a direct result of this study, the following recommendations were made:

(1) Replicate the study. This would provide an opportunity to further validate the findings.

(2) Write definite course objectives that clearly identify the material that the instructor wants the students to learn.

(3) Determine students' attitudes toward evaluation methods used in Economics classes.

The kinds of contextual factors which influence testing and the use of test results are just beginning to be appreciated. The questions of how tests can contribute to teacher understanding of the student, and how to plan a testing program, select tests, and analyze and interpret test results for instructional improvement have been apparent for a number of years (Traxler, 1953). Teacher training, experience and attitudes toward tests and testing seem to be key factors related to test use. Other factors that appear to matter are test-taker characteristics and the options that are available to the teachers.

What research has been done has focused either (1) on procedures for determining reliability of examinee assignments to mastery states (Huynh, 1976; Subkoviak, 1978; Algina and Noe, 1978) or (2) on procedures for the determination of test length that minimizes misclassification errors (Fhanner, 1974; Wilcox, 1977). Rignor and Hambleton (1979) tried to tie these areas together to provide useful results for test practitioners to enable them to determine test lengths to fit the situations in which their tests will be used. Their study concluded that many factors will have an influence on the test length which is finally selected.

Another form of evaluation was that of F. S. Keller who used the principles of reinforcement theory to teach general psychology with student proctors providing the necessary types and amounts of reinforcement. This approach allowed the students to proceed at their own pace. At the

end of each unit, students asked their proctor for a short, ungraded essay test covering the unit's objectives (Gage and Berliner, 1975). Keller's studies have revealed the results for two courses taught one year apart under the strategy was almost identical -- 65% to 70% of the students received A's or B's. Each time the strategy was applied, it produced a large percentage of A's and B's, but few D's and F's (Keller, 1968).

This study provides a base for future studies and experimentation in the area of evaluation methods. It is hoped that new imaginative instructional strategies will more perfectly meet the needs of students in the years ahead.

REFERENCES

Algina, J., and M. J. Noe. A Study of the Accuracy of Subkoviak's Single Administration Estimate of the Coefficient of Agreement using Two True-Score Estimates. Journal of Educational Measurement, 1978.

Baron, D. Evaluation Techniques for Classroom Teachers. New York: McGraw-Hill Book Co., Inc., 1958.

Bauerfeind, R. H. Building a School Testing Program. Boston: Boughton Mifflin, 1963.

Chauncey, Henry, and John E. Dobbin. Testing: Its Place in Education Today. New York: Harper & Row Publishers, Inc., 1963.

Cohen, A. M. "Developing Specialists in Learning." Junior College Journal, September 1972.

Cross, K. P. "New Students of the 70's." The Research Reporter, V. 1 (4) 1-5, 1971.

Fearn-Wannan, H. J. Conclates of Academic Productivity, Unpublished Ph.D. Thesis, Monash University, Melbourne, 1977.

Fearn-Wannan, H. J. Students' Perceptions of Their Lectures as Determinants of Academic Performance. Paper delivered at the annual conference of the Society for Research in Higher Education, Brighton, England, December, 1979.

Fhaner, S. Items Sampling and Decision-Making in Achievement Testing. British Journal of Mathematical and Statistical Psychology, 1974.

Gage, N. L., and D. C. Berliner. Educational Psychology. Rand McNally College Publishing Co., Chicago, IL., 1975.

Gorrow, F. F. Better Classroom Testing. San Francisco: Chandler Publishing Co., 1966.

Hutchins, R. M. The Learning Society. New York: New American Library, 1968.

Huynh, R. On the Reliability of Decisions in Domain-Referenced Testing. Journal of Educational Measurement, 1976.

Keller, F. S. Good-Bye Teacher! Journal of Applied Behavior Analysis, 1968.

Kirkland, M. "The Effects of Tests on Students and Schools." Review of Educational Research, 1971, 303-350.

Kryspin, William J., and John F. Feldhusen. Developing Classroom Tests. Minneapolis: Burgess Publishing House, 1974.

Mink, O. The Behavior Change Process. New York: Harper & Row, 1970.

Quinto, F. "Teacher-Made Tests -- An Alternative to Standardized Tests." Today's Education, 1977.

Rignor, Daniel R., and Ronald K. Hambleton. Effects of Test Length and Advancement Score on Several Criterion-Referenced Test Reliability and Validity Indices. Paper presented at the annual meeting of the National Council on Measurement in Education, 1979.

Roueche, J. E., and John C. Pitman. A Modest Proposal: Students Can Learn. San Francisco: Jossey-Bass Publishers, 1972.

Subkoviak, M. J. The Reliability of Mastery Classification Decisions. Paper presented at the First Annual Johns Hopkins University National Symposium on Educational Research, Washington, October 27, 1978.

Traxler, A. E. Introduction to Testing and the Use of Test Results in Public Schools. Westport, Conn.: Greenwood Press, 1953.

Wilcox, R. R. Estimating the Likelihood of False-Positive and False-Negative Decisions in Mastery Testing: An Empirical Bayes Approach. Journal of Educational Statistics, 1977.

Wilson, John A., Mildred Robeck and William Michael. Psychological Foundations of Learning and Teaching. San Francisco: McGraw-Hill Book Company, 1969.

Yeh, J. P. Test Use in Schools. Los Angeles, California: Center for the Study of Evaluation, 1978.